

Cassette Search

Cassette Search allows searching genes that are part of chromosomal cassettes, which is defined as a stretch of protein coding genes with intergenic distance smaller or equal to 300 base pairs, involving specific protein clusters. First, select the protein cluster underlying the cassettes (COG or Pfam), the protein cluster identifier for the search (cluster/function identifier, cluster/function name, or both), the logical operator (either “and” or “or”) used for the search expression and the order of presenting the search results.

Cassette Search

Max. of 100 genomes can be selected.

Select Protein Cluster	Function Search Field	Logical Operator	Sort Results
<input checked="" type="radio"/> COG	<input checked="" type="radio"/> Function ID	<input checked="" type="radio"/> And (intersection)	<input checked="" type="radio"/> Function matches
<input type="radio"/> Pfam	<input type="radio"/> Function Name	<input type="radio"/> Or (union)	<input type="radio"/> Cassette gene count
	<input type="radio"/> Both (Id and Name)		

Search Text
Use commas to separate search terms.

COG0126, COG0149, COG0166

Sequencing Status
All Finished, Permanent Draft and Draft

☒ List ☐ Tree Show

Search for: <enter a genome name to

- Salinispora tropica CNB536 (B) [P]
- Salinispora tropica CNH898 (B) [P]
- Salinispora tropica CNR699 (B) [P]
- Salinispora tropica CNS197 (B) [P]
- Salinispora tropica CNS416 (B) [P]
- Salinispora tropica CNT250 (B) [P]
- Salinispora tropica CNT261 (B) [P]
- Salinispora tropica CNY012 (B) [P]
- Salinispora tropica CNY678 (B) [D]
- Salinispora tropica CNY691 (B) [D]

Search

Cassette Search Results

Search Terms
COG0126
COG0149
COG0166

Page 1 - 100 shown per page
Sorted by the number of matches

Cassette ID	Cassette Gene Count	Function ID	Function Name	Gene ID	Genome Name
Number of function matches: 3					
251516232000002	36	COG0126	3-phosphoglycerate kinase	2515520224	Salinispora tropica CNH898
		COG0149	Triosephosphate isomerase	2515520223	
		COG0166	Glucose-6-phosphate isomerase	2515520218	
Number of function matches: 3					
251759712400006	42	COG0126	3-phosphoglycerate kinase	2517997237	Salinispora tropica CNS416
		COG0149	Triosephosphate isomerase	2517997238	
		COG0166	Glucose-6-phosphate isomerase	2517997243	
Number of function matches: 3					
254034865500012	43	COG0126	3-phosphoglycerate kinase	2540888139	Salinispora tropica CNT250
		COG0149	Triosephosphate isomerase	2540888138	
		COG0166	Glucose-6-phosphate isomerase	2540888133	

Figure 1. Cassette Search and Search Result.

In the example shown in Figure 1(i), the search involves COG clusters, cluster identifiers are used for the search condition which involves “and”: search genes that are part of cassettes involving COG0126, COG0149 and COG 0166 (which are part of the *Glycolysis pathway*). Users are required to select up to 100 genomes to search. The search can be limited based on domain and sequencing status. In this example, we select all *Salinispora tropica* genomes in IMG. The Cassette Search Result lists the genes that satisfy the search condition, together with the identifiers of the cassettes they are part of, their associated protein cluster identifiers and names, and their genomes, as illustrated in Figure 1(ii).

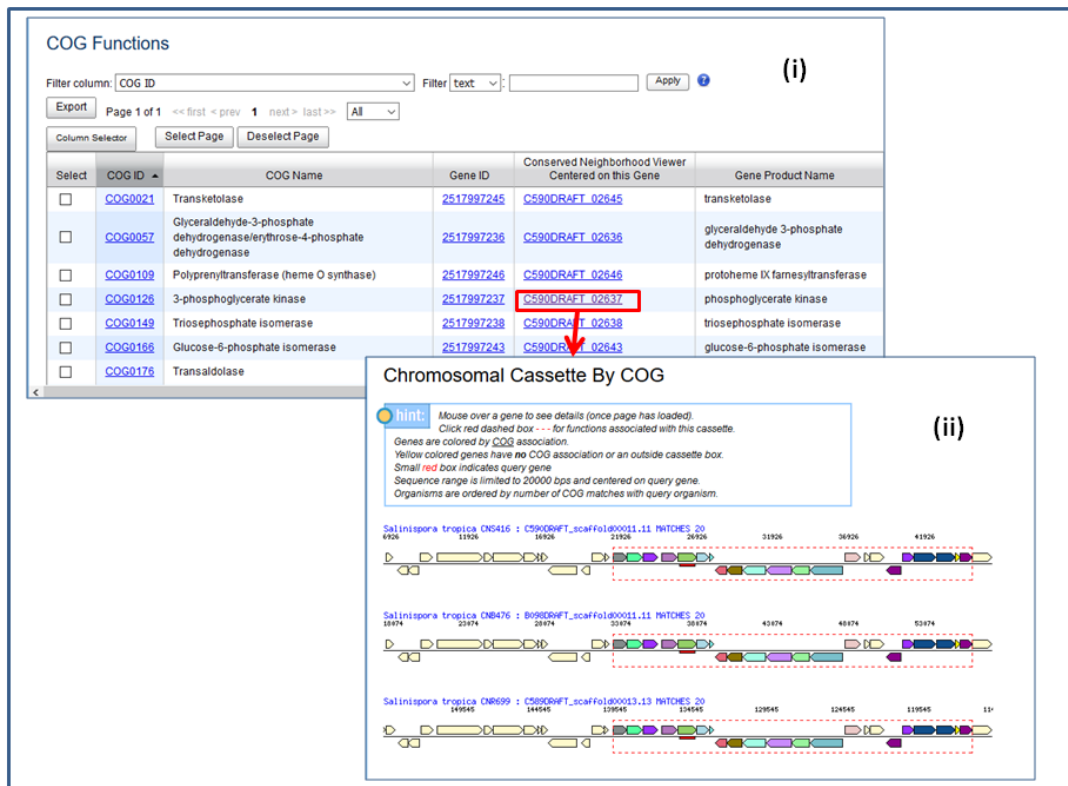


Figure 2. Exploring Cassette Search Result.

The cassette identifiers provide links to the Chromosomal Cassette details page. For example, selecting Cassette ID 251759712400006 (as highlighted in Figure 1(ii)) will lead to the Chromosomal Cassette COG page shown in Figure 2(i). Selecting one of the ID listed in the **Conserved Neighbor Viewer Centered on the Gene** column will lead to the Chromosomal Cassette display as shown in Figure 2(ii).

(Please note that Cassette ID values can change over time. This is not a preserved ID value such as genome or gene IDs.)